

On Measuring Aggregate “Social Efficiency”

Martin Ravallion*

World Bank, 1818 H Street NW, Washington DC

Abstract: Cross-country comparisons of social indicators controlling for income and/or social spending have been widely used to measure and explain “social efficiency,” analogously to “technical efficiency” in production. The paper argues that these methods are clouded in ambiguities about what exactly is being measured. Standard methods of measuring technical efficiency require assumptions that seem unlikely to hold for social indicators. In the context of a simple parametric model of life expectancy, conditions are identified under which there will be a systematic pattern of bias in estimates of efficient health spending.

Keywords: Social indicators; human development; poverty; social spending; efficiency frontiers.

JEL: D61, I12, O57

World Bank Policy Research Working Paper 3166, November 2003

The Policy Research Working Paper Series disseminates the findings of work in progress to encourage the exchange of ideas about development issues. An objective of the series is to get the findings out quickly, even if the presentations are less than fully polished. The papers carry the names of the authors and should be cited accordingly. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the view of the World Bank, its Executive Directors, or the countries they represent. Policy Research Working Papers are available online at <http://econ.worldbank.org>.

* Address for correspondence: mravallion@worldbank.org. For comments the author is grateful to Angus Deaton, Jed Friedman, Aart Kraay, Erwin Tiongson and Adam Wagstaff.

1. Introduction

Invariably many of the things relevant to assessing a county's performance in promoting human development and reducing poverty are not directly observed by stakeholders. For example, governmental efforts in delivering social services to those in most need are not readily observable by the people who finance the spending and/or vote for those responsible. Efforts to make governments more accountable, and make development assistance more performance driven, beg for reliable methods of assessing latent aspects of country performance. This would allow aid donors and domestic tax payers to determine how much social outcomes might be improved by better use of an economy's existing resources.

This paper provides a critical overview of the most common approach found in the literature. By this approach, one attempts to infer the "social efficiency" of an economy from the measured deviations of an observed social indicator — such as average life expectancy, the infant mortality rate or the literacy rate — from an efficiency frontier, typically identified from the residuals of a regression of that indicator on control variables such as mean income and public spending on social services. The econometric tools used have largely been borrowed from the literature on measuring technical efficiency in production.

What can these methods tell us about latent aspects of country performance in improving social outcomes? Naturally there are (potentially important) concerns about the quality of the data. However, the present paper will put such concerns aside; before any method is taken to bigger and better data sets, or even implemented on existing imperfect data sets, we should look closely at its theoretical foundations. Nor does the paper present any new empirical findings. Instead, the sole aim is to critically assess whether existing methods can be expected to reliably enhance public knowledge about an economy's efficiency in achieving agreed social goals.

The following section describes examples of these methods found in the literature. Section 3 then points to a number of concerns about the conceptual foundations and empirical reliability of these methods. Section 4 elaborates on the specific sources of bias in estimates of social efficiency in the context of a simple expository model. Section 5 concludes.

2. Approaches in the literature

In comparing social indicators across countries it has become common to control for income differences. In an early and influential example, Sen (1981) looked at the deviations of actual log life expectancies from their predicted values obtained by regressing on log income per capita. These residuals suggested that Sri Lanka was the best performer amongst developing countries. Sen showed that the predicted national income corresponding to Sri Lanka's (high) life expectancy was about 20 times higher than the country's actual income. This excellent conditional performance was attributed to Sri Lanka's high level of social spending over many decades. By interpretation, Sri Lanka was deemed to be a good performer in human development because a relatively large share of its economic output was devoted to activities that are good for health.

Such conditional comparisons of social indicators and poverty measures across countries have since become common in both the academic literature and policy discussions. They have taken the form of either the "horizontal" comparison made by Sen (in which the difference in performance is measured in the units of the horizontal axis) or the straight "vertical" residuals (in units of the social indicator). One can now find many examples in studies of developing country performance in human development. For example, the WHO's *World Health Report* for 1999 provides health "performance measures" over time by country, based on the residuals from regressing health aggregates on the log of GDP per capita, its squared value and a trend (WHO,

1999, Annex Table 6). The residuals are taken to reveal public-sector performance, notably through the expansion and dissemination of knowledge about health care. It is claimed that some countries have performed considerably better than others, as assessed by this method. Other international agencies have also used residual comparisons; examples can be found in both the World Bank's *World Development Reports* and the UNDP's *Human Development Reports* (see, for example, World Bank, 1993, and UNDP, 1996). Other examples of residual comparisons of country performance include Kakwani (1993), Wang et al., (1999) and Moore et al (1999).

Taking the idea a step further, a number of papers and reports have tried to assess country performance relative to an “efficiency frontier” based on the best performing countries in terms of social indicators or poverty measures conditional on the measured covariates. The latter vary between applications, but typically include one or both of mean income and social spending per capita and possibly other controls. There are numerous examples in the literature, and the following discussion only focuses on four illustrative cases:¹

- In its *World Health Report 2000*, the WHO gives country rankings of the efficiency of national health care spending in raising DALE (“disability-adjusted life expectancy”).² These rankings were based on efficiency frontiers calibrated to regressions of life expectancy on health expenditure (viewed as the “input”) and schooling attainments (to proxy for “non-health system determinants of health”) (Evans et al., 2000).
- In research done in the IMF, Gupta et al., (1997) and Gupta and Verhoeven (2001) assessed the efficiency of government spending on health and education across countries and over time against efficiency frontiers calibrated to data on social indicators for health

¹ Also see Fakin and de Crombrugghe (1997), Wang et al., (1999), Moore et al., (1999), Clements (2002), Afonso et al., (2003), Afonso and St. Aubyn (2003) and Hollingsworth and Wildman (2003).

² For a discussion of DALEs see Anand and Hanson (1997).

and education. They found that countries in Africa are less efficient on average than elsewhere, though Africa's efficiency appears to have improved over time.

- A similar idea has also been used to assess the efficiency of public spending on social services in reducing poverty. Gouyette and Pestieau (1999) regressed measures of poverty and inequality on levels of social spending across OECD countries and use the residuals to construct an efficiency frontier, identifying Belgium as the benchmark country, with lowest poverty given its social spending.³
- In research at the World Bank, Jayasuriya and Wodon (2003) derived measures of the efficiency of countries (and provincial governments within countries) in attaining the Millennium Development Goals. Their frontier was based on regressions of social indicators on mean income and social spending. On this basis, they argued that substantial progress is possible through more efficient use of existing resources.

The methods of fitting an efficiency frontier found in this literature have all been borrowed from the literature on the measurement of technical inefficiency in production. In estimating production functions, one can allow the possibility that there is technical inefficiency such that actual output is less than the maximum output obtainable at given inputs. Various methods have been used to determine the efficiency frontier. Sometimes it is fitted non-parametrically to the observations with the best measured performance at each input level. Thus the frontier is the upper boundary of the smallest set encompassing the data points on outputs and inputs. An example of this approach is the “free disposal hull” (FDH) method used (in the context of measuring social efficiency) by Fakin and de Crombrugghe (1997) and Gupta and Verhoeven (2001). In other applications, a parametric model of the social indicator as a function

³ For further discussion of the results of Gouyette and Pestieau (1999) see Ravallion (2001).

of postulated covariates has been used to identify the frontier for social indicators. The examples in Gouyette and Pestieau (1999), WHO (2000) and Evans et al., (2000) use variations on one of the oldest methods used in production analysis to estimate the extent of technical inefficiency, sometimes called the Corrected Ordinary Least Squares (COLS) (Kumbhakar and Lovell, 2000). The production function is first estimated by regressing output on a vector of inputs, and then the intercept is shifted upwards such that the production frontier bounds the data from above, i.e., by finding the largest (positive) residual.⁴ With panel data one can implement a variation on this method in whereby the frontier is anchored to the country with the maximum intercept in a country fixed effects model estimated on panel data (following the method proposed by Schmidt and Sickles, 1984, for measuring technical efficiency in production using panel data). This is the method used by Evans et al., (2000) to quantify their frontier in assessing the comparative efficiency of national health systems.

An alternative to COLS-type methods is the stochastic frontier (SF) production function whereby the error term in the regression for output includes a one-sided component (representing inefficiency) as well as a regular zero mean error (Aigner et al., 1977; Meeusen and van den Broeck, 1977). The SF method has the advantage over COLS and the main competing non-parametric methods that it allows for random deviations from the frontier (in both directions), such as due to measurement errors or shocks. Thus, not all of the data need be in the production set. In estimating the model's parameters it is assumed that both error terms are independent and identically distributed, and it is usually assumed that the zero-mean component is normally distributed while inefficiency component is half normal. This is the method for measuring efficiency in reaching human development goals used by Jayasuriya and Wodon (2003).

⁴ This method appears to have been first proposed by Winsten (1957) in his comments on Farrell (1957). Subsequent variations on this method are discussed in Kumbhakar and Lovell (2000).

The literature has not stopped at measuring social efficiency, but has tried to explain the revealed differences across countries or provinces. Differences in social policies are one possible explanation. This was how Sen (1981) explained why Sri Lanka appeared to be an outstanding performer in human development given its income, although this explanation was the subject of subsequent debate (Bhalla and Glewwe, 1986; Sen, 1988; Anand and Ravallion, 1993; Aturupane et al., 1994). Of course, differences in social policies are not the only explanation that can be offered. Another source of heterogeneity in social outcomes at given mean income is cross-country differences in the distribution of income. It has been argued that aggregate health indicators such as life expectancy and the infant mortality rate depend far more on incomes of the poor than the nonpoor (Bidani and Ravallion, 1997). Then the incidence of income poverty also matters to human development, independently of social policies.

There have also been attempts to explain the measured differences in “social efficiency” using a second-stage regression as a means of sorting out which of the possible sources of these differences matters most. This entails regressing the efficiency measure derived from the first stage on other variables. For example, Moore et al., (2000) measure “efficiency in converting national material resources into human development” by the residuals from a regression of a human development index on income. They then regress this efficiency measure on a set of explanatory variables, including a measure of the quality of government institutions. Similarly, Jayasuriya and Wodon (2003) retrieve a measure of inefficiency in attaining human development goals (using the SF specification for production functions) and then regress this measure on indicators of the quality of government and urbanization. They find evidence of an inverted-U relationship with urbanization (and governance, though less markedly), whereby social

efficiency first increases as developing countries urbanize but then starts to decline at sufficiently high levels of urbanization.

3. Pitfalls in measuring and explaining social efficiency

It would be an impressive achievement — indeed a remarkable one — to extract credible measures of latent inefficiencies in attaining social objectives from the type of aggregate country-level data used in this literature. But has that really been achieved?

A problem in assessing these methods is that the theoretical foundation for the empirical models of social indicators has never been clear. We are told that they should be interpreted as empirical “production functions.” A standard assumption in the economic theory of production is free disposability, meaning that if the point (x, y) , for an output y and inputs x , is in the producer’s production set then so too is any point (x', y') such that $x' \geq x$ and $y' \leq y$. As noted in the last section, the assumption of free disposability has been invoked explicitly in some studies of social efficiency and is implicit in other studies. This may be a defensible assumption for a production process (though it can certainly be questioned in that context). But how can we interpret the application of this assumption to (say) life expectancy as the “output” and public spending on health as the “input”? There are (thankfully) very few governments in the world that can freely dispose of their citizens such that if the country initially has a life expectancy of (say) 60 years, and health spending of (say) \$100 per person per year, it is equally feasible for it to have a life expectancy of 40 at the same or greater spending.

The applicability of production theory to measuring social efficiency is questionable. Social indicators do not stem from anything one could reasonably think of as a production function representing a well-defined technology operated by an individual producer with well-defined physical inputs. While there are production functions under the surface somewhere,

there is clearly a lot more going on in determining the aggregate relationship between measured social outcomes and social spending and/or national income.

Without specifying a complete model it is hard to assess the specification choices made in this literature. But there is already enough to make one skeptical. The accounting of “outputs” is worryingly incomplete, such as by focusing on health only, or just one aspect of health. This raises the concern that public spending that is deemed to be inefficient with respect to the partial social indicator may be of value with respect to some omitted indicator. And even if one accepts that life expectancy (say) is an adequate proxy for the “outputs” of the “health system,” the accounting of “inputs” is rarely convincing. For example, the practice in the literature of using public spending on health (say) as a measure of the inputs to health production is hard to defend on theoretical grounds; if anything, one would be more tempted to interpret these regressions as some kind of inverse cost function, in which the cost of an entire bundle of inputs depends on the output level. However, the input prices would then have to be included for a correctly specified empirical model. The omission of these (country-specific) prices means that what is being called “inefficiency” may be nothing more than how public health authorities respond to the input prices they face, including wages.

Similarly, why should one not also control for other types of public spending, recognizing that there is rarely a clear one-to-one mapping between types of spending and measured social indicators? Public spending on health care is not just about raising life expectancy (say), but is also about improving the quality of peoples’ lives. And public spending on (say) education or infrastructure may well matter greatly to health outcomes.⁵ Similarly, why

⁵ For example, Jalan and Ravallion (2003) find that access to water infrastructure improves child health outcomes in rural India, and that the extent of this effect depends on maternal education and household incomes.

not also control for factors such as distribution of income, or the administrative capabilities of its government? What is being identified as “social inefficiency” in this literature could well stem entirely from omitted interdependencies between types of spending and other country circumstances, combined with a partial accounting of social outcomes.

Nor is it clear that the residuals used to assess efficiency in many of these methods are based on reliable measures of the expected values of the social outcomes conditional on spending, income, or whatever else one chooses to control for. In the context of parametric methods, misspecification of the functional form is known to be a concern in measuring technical efficiency in production. For example, Giannakas et al., (2003) provide Monte Carlo simulations indicating the potential for sizeable bias in the mean efficiency measures from SF methods of estimating production functions due to misspecification of the functional form of the production frontier. They demonstrate that the method can suggest quite high levels of inefficiency (10-30% of output) for fully efficient producers.

Non-parametric methods of setting the frontier can avoid such problems, though they introduce new concerns. One must assume a continuous frontier, which must be interpolated from the discrete data points. Sensitivity to outliers, or to “holes” in the support provided by data, can be expected. Recent advances in nonparametric frontier estimation offer the promise of results that are more robust to outliers and noise, by nonparametrically smoothing the frontier, allowing some data points to be outside the production set (Cazals et al., 2002).

Naturally, the more data one has, the more believable these nonparametric methods become. Micro applications in production analysis often use samples of many thousands of producers. However, in the applications to measuring social efficiency at country-level, one is fitting a continuous frontier to at most 200 data points; indeed the application by Gupta and

Verhoeven (2001) of the FDH method to measuring social efficiency uses data for 37 countries. Simulations by Park et al., (2000) suggest considerable imprecision in FDH estimates of efficiency with sample sizes of 100 or less, even when one is allowing for just a few inputs; the imprecision naturally rises with the number of inputs and falls with the number of data points. Asymptotic results for drawing statistical inferences about estimates of efficiency using the FDH method are now available in the literature (Park et al., 2000), though there do not appear to have been any applications to the measurement of social efficiency.

There are also concerns about the functional forms used in these methods. Given the bounded nature of most social indicators, the linear and even linear-in-logs specifications favored in most empirical work using parametric social indicator regressions cannot possibly be right, at least globally.⁶ The regression residuals will then be some higher-order (nonlinear) function of income and so lose their interpretation as the deviations of actual social outcomes from their expected values conditional on income.

Further concerns arise about whether conditional cross-country comparisons should be based on the levels of the social indicators and the control variables (as in most of the literature) or their changes over time (as advocated by Bhalla and Glewwe, 1986, and Aturupane et al., 1994). Taking differences over time (or deviations from time means) has the usual advantage that country fixed effects correlated with the regressors (that would otherwise bias the results) can be swept away. However, it also raises well-known concerns that the measured changes may not properly capture the effects of interest. One way this can happen is that the changes over time are measured with far greater error than the levels, so that the signal-to-noise ratio deteriorates substantially, with a corresponding increase in bias due to measurement error in the

⁶ Better transformations have been proposed by Anand and Ravallion (1993) and Kakwani (1993).

regressors and higher standard errors. Another way that the “change-on-change” regression of social indicators on income (and possibly other controls) may miss the effects of interest is that the time period is too short (even if there is no measurement error within the chosen period). For example, in response to the claim by Bhalla and Glewwe (1986) that Sri Lanka’s improvement in social indicators was not unusually good relative to its gain in income, Sen (1988) argued that effects of interest largely predated the time period used by Bhalla and Glewwe.⁷

Even putting all these problems to one side, there remain important concerns about the validity of the assumptions made about the error term in current practices for measuring social efficiency by parametric methods using social indicator regressions. One concern is about the distributional assumptions made in the SF method. Any non-normality (especially skewness) in the zero-mean random error component will be incorrectly attributed to inefficiency (as pointed out by Skinner, 1994). And even if normality is deemed to be an acceptable assumption for the zero-mean error component, this is far from clear for the inefficiency component, and it is known that the results obtained can be quite sensitive to relaxing this distributional assumption.⁸

Some of these concerns are known from the literature on measuring technical efficiency in production. However, there is a further problem that is intrinsic to assessments of social efficiency by parametric methods but appears to have been entirely ignored in that literature.⁹ This relates to the validity of a key assumption in all versions of these methods found in practice, namely the assumption that the error component deemed to reflect “inefficiency” is uncorrelated

⁷ Time series data for Sri Lanka are consistent with Sen’s conclusion that a significant and quantitatively important role was played by Sri Lanka’s social spending in reducing infant mortality at given average income (Anand and Ravallion, 1993).

⁸ See, for example, Baccouche and Kouki (2003). Also see the discussion in Greene (1999).

⁹ For example, although the WHO’s health efficiency estimates using the frontier method have attracted a good deal of criticism, the WHO’s (2001, Chapter 11) survey of those criticisms does not mention the following concern.

with the observed control variables. This assumption will be more defensible in some applications than others. When estimating a production function one might be willing to treat the extent of technical inefficiency as being uncorrelated with factor inputs. This is justified under the assumption that the inefficiency is unknown to the producer, and so could not affect input choices. This is a special case of the longstanding argument for using Ordinary Least Squares (OLS) to estimate a production function under the assumption that the production error term is unknown to producers *ex ante*, at the time input choices were made.¹⁰ That assumption is questionable and there is a literature that has attempted to relax it, such as by using longitudinal (panel data) following Mundlak (1963).¹¹ However, one can at least point to a story as to why production inputs might be safely treated as exogenous to technical inefficiency, and this has been the maintained assumption in the literature on frontier production functions.¹²

When modeling social outcomes there must, however, be a reasonable presumption that the error component intended to capture “inefficiency” is correlated with the regressors, both in cross-sectional data and over time. The inefficiency in attaining desired social outcomes presumably stems from social or economic activities that are unproductive from the point of view of those outcomes. Examples include public spending policies that do little or nothing to improve social outcomes and income gains to the rich, which arguably do little to improve attainments in basic health and education or to reduce poverty. Only under rather special conditions will these inefficient components be uncorrelated with total income or public spending. As an economy grows one expects at least some of the income gains to be inefficient

¹⁰ For an interesting recent discussion of how stochastic terms arise in estimating agricultural production functions see Pope and Just (2003).

¹¹ See Olley and Pakes (1996) for a more general treatment, allowing for endogenous exist of producers as well as endogenous input choice.

¹² The assumption is partially relaxed in the method of estimating production frontiers using panel data proposed by Schmidt and Sickles (1984).

from the point of view of certain social goals. Similarly, at least some of an increment to total social spending can be expected to be inefficient. Thus a positive correlation between the level of inefficient income or spending and the totals can be postulated. Standard econometric methods for estimating parametric production frontiers are not then valid for the problem of assessing social efficiency and it is not clear what meaning can be given to the resulting measures.

There are a number of further specification issues in the subset of the literature that has also tried to explain measured differences in social efficiency. Without greater conceptual clarity about what “social efficiency” means in this context, it is difficult to assess the empirical specifications used in this strand of the literature. When estimating a production function it is reasonably clear what variables qualify for the first stage regression — they should be the factor inputs to production.¹³ However, when the same tools are applied to human development it is unclear which variables should be in the first stage (used to measure inefficiencies) and which should be in the second (used to explain inefficiencies). Why, for example, does urbanization only matter to the extent of “social inefficiency” (as in Jayasuriya and Wodon, 2003)?

Urbanization influences the costs of public service provision, which would presumable matter to the efficient level of outcomes too. And why is it deemed “inefficient” for an economy with weak administrative capabilities to devote fewer resources to activities that are intensive in those capabilities? In some papers social spending appears in the second stage (as in Sen’s, 1981, explanation of differences in performance at given incomes, though Sen did not have a second stage regression as such) and sometimes in the first stage (as in Jayasuriya and Wodon, 2003).

¹³ Although the ambiguity about specification choices also arises in the production function literature, since the variables used to in the second stage to explain measured technical efficiency could equally well also qualify as shift parameters in the production frontier. For further discussion see Kumbhakar and Lovell (2000, Chapter 7).

Misspecifications in the first stage will clearly also contaminate the second stage. For example, in their first stage regression, Jayasuriya and Wodon (2003) assume a linear relationship between their social indicators and income. In the second stage, they then find an inverted-U relationship between their efficiency measure and urbanization. However, this could simply reflect a first-stage misspecification, given that the relationship between social indicators and income cannot possibly be linear but is very likely to be concave (given the bounded nature of the social indicators). Even if there is no real effect (linear or otherwise) of urbanization on social efficiency, the efficiency measure will be found to have an inverted-U relationship with urbanization, but only because urbanization acts as a proxy for mean income.

To give another example, it is plausible that differences in the distribution of income will matter to social outcomes at given mean income. Yet, I do not know of any study of social efficiency that has controlled for income distribution. The second-stage covariates of measured social efficiency could then be solely picking up covariates of inequality.

The general point here is that it is unclear what the second stage regression could ever meaningfully tell us about the determinants of “social efficiency” when the measure of the latter is biased and inconsistent because the relevant error component is not in fact orthogonal to the regressors at the first stage. All one might be picking up in the second stage are correlations with the biases passed on from the first-stage.

When monitoring country performance over time, a further concern arises from the fact that there are two distinct sources of the measured changes in social efficiency. Firstly, there may be an unconditional change in the social indicator and secondly there may be a change in the conditioning variable(s). Either could account for the measured improvement or worsening over time in social efficiency. For example, a country that appears to be performing poorly now

in terms of its average health attainments given its income could become a star performer in the future simply by (suffering) negative growth, without any gain in actual health attainments. This is a nagging concern about the assessments of progress in human development using these methods.

4. Sources of bias in a simple expository model

The literature has tended to justify these methods on casual, intuitive grounds, without rigorously defining the theoretical concept one is trying to measure or the conditions under which the methods used will give reliable results. Established methods from the analysis of production functions have been applied to the problem of explaining human development or poverty without due consideration as to whether the methods are appropriate.

This section will try to throw further light on the specific sources of bias in assessments of social efficiency using these methods. Attention is confined here to parametric frontier methods, notably the COLS method, though I will note some implications for the SF method when applied to measuring social efficiency. The analysis will focus on the case in which one is assessing the efficiency of health spending in raising life expectancy, though equally well one might be using this method to assess the efficiency of other types of public spending or, indeed, the efficiency of the economy as a whole (in which case the control variable is national income) with respect to one or more social indicators or aggregate welfare measures. By using only one “input” it will be possible to demonstrate the key points with nothing more than some simple algebra. However, the main messages also apply to versions of this method that add more control variables.

One must first be more precise about what we mean by “social inefficiency.” The definition that appears to be closest to that underlying much of the applied work reviewed in the

previous sections is that social inefficiency refers to that share of spending that is devoted to things that do nothing for a specific social outcome. So total health spending H has two components, one which raises life expectancy and one which does not:

$$H_i = H_i^E + H_i^I \quad (i=1, \dots, n) \quad (1)$$

where $H_i^E \geq 0$ is the efficient component, that is health promoting, while $H_i^I \geq 0$ is the inefficient component. By definition, life expectancy (denoted L_i , which may be some appropriate nonlinear function of actual life expectancy) depends on the efficient component:

$$L_i = \alpha + \beta H_i^E + \varepsilon_i \quad (2)$$

where α and β are parameters and ε_i is a zero-mean i.i.d. error term. However, equation (2) is not estimable since the efficient component is unobserved. The model linking the observed total spending on health to life expectancy is:

$$L_i = \alpha + \beta H_i + \mu_i \quad (3)$$

where $\mu_i = -\beta H_i^I + \varepsilon_i$. It is readily verified that the OLS regression coefficient of L on H is

$\hat{\beta} = \beta \hat{\gamma}$ where $\hat{\gamma} = \text{cov}(H, H^E) / \text{var}(H)$ is the OLS regression coefficient of H^E on H . Since the estimates of social efficiency using the COLS method are based on the OLS regression in (3), and this is biased for all $\hat{\gamma} \neq 1$, all estimates of social efficiency at the country level will also be biased.

What can we say about the direction of bias? Two seemingly natural assumptions are:

Assumption 1: Higher levels of efficient health spending raise life expectancy ($\beta > 0$).

Assumption 2: Both the efficient and inefficient components of spending tend to rise with total spending ($0 < \hat{\gamma} < 1$).

Under these assumptions, the OLS estimate of the regression coefficient of life expectancy on health spending will be positive but will underestimate the true impact of efficient health spending on life expectancy.

The same source of bias is found in the SF method, in which one allows explicitly for a one-sided error component representing “inefficiency.” To see why, re-write equations (3) as:

$$L_i = [\alpha - \beta E(H_i^I)] + \beta H_i + [\mu_i + \beta E(H_i^I)] \quad (4)$$

Given that ε_i is a zero-mean error term, the transformed error term in (4) now has zero mean.

So one is tempted to estimate β by OLS. One can then retrieve estimates of the other parameters (α and the variances of H_i^I and ε_i) by invoking the distributional assumptions in the SF model (i.e., that ε_i is normally distributed and H_i^I is positive half-normal).¹⁴ Thus we appear to have everything needed to measure social efficiency at country level.

However, all this breaks down as soon as one recognizes that applying OLS to equation (4) does not give a consistent estimate of β for the reason discussed above, namely that $\text{cov}(H_i, \mu_i) \neq 0$ in general. Again, the bias in estimating the slope of the frontier is passed onto the estimates of social efficiency.

What are the implications for estimates of the social efficiency of spending? Recall that by the COLS method one estimates the level of socially efficient spending in each country by shifting the intercept of the social indicator regression until it passes through the data point for the country with the largest (negative) residual. The equation of this frontier is thus

$$L_i = \hat{\alpha} + \hat{\beta} \hat{H}_i^E + \hat{\mu}_i^* \text{ where } \hat{\alpha} \text{ and } \hat{\beta} \text{ are the OLS estimates of the parameters of (3) and}$$

¹⁴ For the present purpose, this two step method-of-moments procedure is equivalent to using maximum likelihood in one step; for further discussion of the two-step method see Kumbhakar and Lovell (2000, Chapter 3).

$\hat{\mu}^* = \max(\hat{\mu}_1, \dots, \hat{\mu}_n)$ where $\hat{\mu}_i = L_i - \hat{\alpha} - \hat{\beta} H_i$. An estimate of the health-promoting output for country i (\hat{H}_i^E) is then obtained by inverting the equation for the frontier, giving:

$$\hat{H}_i^E = \frac{L_i - \hat{\alpha} - \hat{\mu}^*}{\hat{\beta}} = \hat{H}^* + \frac{L_i - \hat{L}^*}{\hat{\beta}} \quad (5)$$

(noting that $\hat{L}^* = \hat{\alpha} + \hat{\beta} \hat{H}^* + \hat{\mu}^*$). Notice, however, that this is nothing more than a fixed linear transformation of the observed life expectancy for country i . If the aim is simply to rank countries by their socially efficient spending then this method is not telling us anything more about any specific country than we already know from the observed (unconditional) life expectancy.

How does equation (5) compare to the true social efficiency of spending in country i ?

Inverting equation (2) we find that:

$$H_i^E = H^* + \frac{L_i - L^* + \varepsilon^* - \varepsilon_i}{\beta} \quad (6)$$

where (H^*, L^*) is the data point for the true benchmark country, for which the error term is

$\mu^* = \varepsilon^* = L^* - \alpha - \beta H^* = \max(\mu_1, \dots, \mu_n)$, where $\mu_i = L_i - \alpha - \beta H_i$. Subtracting equation (6)

from (5) and re-arranging terms it is readily verified that:

$$\hat{H}_i^E - H_i^E = \frac{1}{\beta} \left[(\mu^* - \hat{\mu}^{**}) + (L_i - \hat{L}^*) \left(\frac{1}{\hat{\gamma}} - 1 \right) + (\varepsilon_i - \varepsilon^*) \right] \quad (7)$$

where $\hat{\mu}^{**} \equiv \hat{L}^* - \alpha - \beta \hat{H}^*$. The first term in square brackets on the RHS of equation (7),

$\mu^* - \hat{\mu}^{**} = L^* - \hat{L}^* - \beta(H^* - \hat{H}^*)$, measures the extent to which the true residual of the

benchmark country is underestimated when evaluated at the true parameters. It can be readily

verified that if the data form a convex set then $\mu^* \geq \hat{\mu}^{**}$ (with the inequality strict if the set is

strictly convex), so this term would be a source of upward bias to estimates of the efficient level of spending. One might be happy to make such a convexity assumption if this was a production set. But that is not the case. The discreteness of the data alone will generate non-convexities. More generally, there appears to be little one can say on *a priori* grounds about the sign of this first term, as its value will be data-specific.¹⁵ I will set this term to zero under the following (admittedly *ad hoc*) assumption:

Assumption 3: The benchmark country is a sufficient outlier that it stays being the country with highest conditional life expectancy after correcting for the bias in the slope of the efficiency frontier, i.e., $\mu^* = \hat{\mu}^{**}$.

The second term in square brackets in (7) arises from the bias in estimating the slope of the frontier, as already discussed. This term will impart a downward (upward) bias to estimates of social efficiency for all countries with life expectancy less than (greater than) that for the country with the highest conditional life expectancy.

The third term in squared brackets ($\varepsilon_i - \varepsilon^*$) reflects country-specific heterogeneity. Even at mean points ($E(\varepsilon) = 0$), unobserved variables that influence life expectancy in the benchmark country at given health spending will lead the method to miss-identify the vertical location of the frontier. (Under the distributional assumptions about the error terms in the SF method, this effect will vanish in expectation. In panel data versions of the COLS method, only the time invariant component of this heterogeneity term matters.)

We have seen that the direction of bias is hard to predict at the level of individual countries. What can we say about average social efficiency? Under Assumptions 1-3, it is

¹⁵ Given that β is underestimated, it must be the case that $L^* < \hat{L}^*$ and that $H^* < \hat{H}^*$; however, this is not sufficient for determining the sign of $\mu^* - \hat{\mu}^{**}$ given that $\beta > 0$.

readily verified that the asymptotic bias in the estimate of mean socially efficient spending is given by:

$$\text{Bias} \equiv \text{plim}(\sum \hat{H}_i^E / n) - E(H^E) = \frac{1}{\beta} \left[(\bar{L} - L^*) \left(\frac{1}{\hat{\gamma}} - 1 \right) - \varepsilon^* \right] \quad (8)$$

Two sources of asymptotic bias are now evident in the squared brackets on the RHS of (8). The first source is the bias in the slope of the frontier arising from the correlation between inefficient spending and total spending while the second is the bias in the height of the frontier arising from latent heterogeneity in the benchmark country.

The history of social policy in the frontier country is clearly a potentially important source of latent heterogeneity in conditional social outcomes. If the country with the best conditional performance got to that point by a long history of favorable policies (and not just for health) then life expectancy will tend to be higher than one would expect given current health social spending. From equation (8) we can see that if the benchmark country has above average life expectancy ($\bar{L} < L^*$) as well as favorable latent conditions for life expectancy ($\varepsilon^* > 0$) then mean social efficiency will be underestimated.

Under certain conditions it is possible to use an instrumental variables estimator to correct the bias in the original social indicator regression. The practical challenge would be to find an instrumental variable (IV) that is correlated with the socially efficient component of spending but uncorrelated with the inefficient component. It is far from obvious that such an IV exists, though this route may merit further research, such as by using lagged observations over time as IV's in a panel-data structure. Notice, however, that it cannot be presumed that simply correcting for the bias in the regression coefficient on spending will reduce the overall bias in the

estimate of mean social efficiency. That would only hold if the two sources of bias discussed above work in the same direction.

5. Conclusions

A strand of the literature on human development and poverty has applied existing tools for measuring technical efficiency in production to the problem of measuring aggregate “social efficiency.” The paper has questioned whether these methods can deliver credible results.

There is a nagging concern that what is being called “inefficiency” in this literature may reflect nothing more than how arbitrarily omitted differences in country circumstances — such as differences in the prices faced, or other relevant types of public spending, or administrative capabilities — influence partially measured social outcomes. The set of feasible combinations of social outcomes and levels of income and social spending in any economy is almost certainly riddled with non-convexities (“holes”) arising from real constraints on what governments can and cannot do. Without specifying which of those constraints is deemed to be binding in assessing “social efficiency” and which is not, it is difficult to make sense of the calculations.

Even if one accepts the free disposability assumption for social outcomes, there are some poorly resolved specification issues for these “social indicator production functions.” In contrast to production analysis (for which it is reasonably clear what constitutes a production input) it is not clear what should be a control variable in measuring social efficiency and what should be used to explain inefficiency. This throws doubt on both the measures obtained, and the explanations that have been given in the literature for measured differences in efficiency across countries.

However, even putting these concerns aside, there are other reasons to question the reliability of these estimates. The main reason for “inefficiency” in attaining desired social

outcomes in human development and poverty reduction is presumably that there are public and private activities that do not promote those goals. Obvious candidates include inefficient social policies (that do not reach those in most need due to insufficient outlays or design deficiencies) and persistently high levels of income inequality (whereby a large share of the aggregate income differences between countries arguably does little for human development or poverty reduction). These sources of “social inefficiency” are typically buried in the regression error term for the social indicator. Even with unbiased estimates of these error terms, disentangling the inefficiency from other factors (including measurement errors) is clearly problematic. However, the least-squares regression parameters and (hence) the residuals can be expected to be biased in general. This arises when total social spending (or national income) is correlated with its own components, including both the efficient and inefficient activities from the point of view of social outcomes. This source of bias has been routinely ignored in the literature on social efficiency.

We have seen that a systematic pattern of downward bias in measures of aggregate social efficiency can be expected under certain conditions that are not obviously implausible. In particular, mean social efficiency will be underestimated by standard parametric frontier methods as long as both the efficient and inefficient components rise with total spending, the best conditional performer is a sufficient “outlier” in the data, and the frontier country or countries tend to have above average (unconditional) performance — possibly reflecting favorable latent conditions for human development stemming from a past history of good social policies and/or low inequality. However, overestimation of social efficiency in certain countries, and even at the mean, cannot be easily ruled out.

These observations point to serious limitations of past attempts to measure and explain social efficiency using cross-country comparisons of observed aggregate data. It is not clear what can be inferred about average efficiency using these methods, even in large samples. And it is problematic indeed to use this type of method to assess and monitor the performance of any specific country, or to explain cross-country differences in performance.

Some of the concerns raised in this paper relate solely to the parametric methods based on social indicator regressions. Nonparametric methods can avoid some of these problems and this may be a more promising route, though the “curse of dimensionality” comes to the fore in applications on cross-country data sets.

However, greater conceptual clarity about the definition and origin of “social inefficiency” is begging before applied work on this topic borrows even more sophisticated tools from production analysis. Whether some other approach to measuring “social efficiency” can yield more credible results remains an open question. There does appear to be scope for less inferentially ambitious approaches based on information pertaining more directly to sources of low performance in reaching agreed social goals in specific settings, including for specific public programs.

References

- Aigner, D., C. Lovell and P. Schmidt (1977). Formation and estimation of stochastic frontier production models, *Journal of Econometrics*, 6, 21-37.
- Afonso, Antonio and Miguel St. Aubyn (2003). Non-Parametric Approaches to Education and Health Expenditure Efficiency in the OECD, mimeo, Technical University of Lisbon, Lisbon, Portugal.
- Afonso, Antonio, Ludger Schuknecht and Vito Tanzi (2003). Public sector efficiency: An international comparison, European Central Bank Working Paper No. 242, European Central Bank, Frankfurt, Germany.
- Anand, Sudhir and K. Hanson (1997). Disability-adjusted life years: A critical review. *Journal of Health Economics* 16, 685-702.
- Anand, Sudhir and Martin Ravallion (1993). Human development in poor countries: On the role of private incomes and public services, *Journal of Economic Perspectives*, 7, 133-150.
- Aturupane, Harsha, Paul Glewwe and Paul Isenman (1994). Poverty, human development and growth: An emerging consensus? *American Economic Review, Papers and Proceedings*, 84(2), 244-249.
- Bhalla, Surjit S., and Paul Glewwe (1986). Growth and equity in developing countries: A reinterpretation of the Sri Lankan experience, *World Bank Economic Review*, 1, 35-63.
- Bidani, Benu and Martin Ravallion (1997). Decomposing social indicators using distributional data, *Journal of Econometrics*, 77(1), 125-140.
- Baccouche, Rafik and Mokhtar, Kouki (2003). Stochastic production frontier and technical inefficiency: A sensitivity analysis, *Econometric Reviews*, 22(1), 79-91.

- Cazals, Catherine, Jean-Pierre Florens and Leopold Simar (2002). Nonparametric frontier estimation: A robust approach, *Journal of Econometrics* 106, 1-25.
- Clements, Benedict (2002). How efficient is education spending in Europe? *European Review of Economics and Finance* 1(1), 3-26.
- Evans, David B., Ajay Tandon, Christopher J.L. Murray and Jeremy A. Lauer (2000). The comparative efficiency of national health systems in producing health: An analysis of 191 countries. GPE Discussion Paper 29, World Health Organization, Geneva.
- Fakin, Barbara and Alain de Crombrughe (1997). Fiscal adjustment in transition economies: Social transfers and the efficiency of public spending. Policy Research Working Paper 1803, World Bank, Washington DC. <http://econ.worldbank.org/resource.php?type=5>
- Farrell, M.J. (1957). The measurement of productive efficiency, *Journal of the Royal Statistical Society A*, 120(3), 253-281.
- Giannakas, Konstantinos, Kien C. Tran and Vangelis Tzouvelekas (2003). Predicting technical efficiency in stochastic production frontier models in the presence of misspecification: A Monte-Carlo analysis. *Applied Economics* 35, 153-161.
- Gouyette, Claudine and Pierre Pestieau (1999). Efficiency of the welfare state, *Kyklos*, 52, 537-553.
- Greene, W.H. (1999). Frontier production functions, in *Handbook of Applied Econometrics*, edited by M.H. Pesaran and P. Schmidt, Oxford: Blackwell Publishers.
- Gupta, Sanjeev, K. Honjo and Martijn Verhoeven (1997). The efficiency of government expenditure: Experience from Africa, Working Paper 97/153, International Monetary Fund. <http://www.imf.org/external/pubs/cat/longres.cfm?sk=2409.0>

- Gupta, Sanjeev and Martijn Verhoeven (2001). The efficiency of government expenditure: Experience from Africa, *Journal of Policy Modeling* 23, 433-467.
- Hollingsworth, Bruce and John Wildman (2003). The efficiency of health production: Re-estimating the WHO panel data using parametric and non-parametric approaches to provide additional information, *Health Economics* 12, 493-504.
- Jalan, Jyotsna and Martin Ravallion (2003). Does piped water reduce diarrhea for children in Rural India? *Journal of Econometrics* 112, 153-173.
- Jamison, D.T., J.Wang, K. Hill and J.L. Londono (1996). Income, mortality and fertility in Latin America: Country level performance, 1960-90, *Revista-de-Analisis-Economico*, 11(2), 219-61.
- Jayasuriya, Ruwan and Quentin Wodon (2003). *Efficiency in Reaching the Millennium Development Goals*, World Bank, Washington DC.
http://publications.worldbank.org/ecommerce/catalog/product?item_id=2435559
- Kakwani, Nanak (1993), Performance in living standards: An international comparison, *Journal of Development Economics*, 41, 307-336.
- Kumbhakar, Subal C., and C.A. Knox Lovell (2000). *Stochastic Frontier Analysis*, Cambridge: Cambridge University Press.
- Moore, Mick, Jennifer Leavy, Peter Houtzager and Howard White (2000). *Polity Qualities: How Governance Affects Poverty*, Working Paper 99, Institute of Development Studies, University of Sussex. <http://www.ids.ac.uk/ids/bookshop/wp/wp99.pdf>
- Meeusen, W., and J. van den Broeck (1977). Efficiency estimation from Cobb-Douglas production functions with composed error. *International Economic Review* 18(2), 435-444.

- Mundlak, Y. (1963). Estimation of production and behavioral functions from a combination of cross-section and time series data, in C. Christ et al. (eds) *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Stanford: Stanford University Press.
- Olley, G. Steven and Ariel Pakes (1996). The dynamics of productivity in the telecommunications equipment industry, *Econometrica* 64(6), 1263-1297.
- Park, B.U., L. Simar and Ch. Weiner (2000). The FDH estimator for productivity efficiency scores, *Econometric Theory* 16, 855-877.
- Pope, Rulon D., and Richard E. Just (2003). Distinguishing errors in measurement from errors in optimization, *American Journal of Agricultural Economics*, 85(2), 348-358.
- Ravallion, Martin (2001). On assessing the efficiency of the welfare state: A comment, *Kyklos* 54(1), 115-123.
- Schmidt, Peter and R.C. Sickles, (1984). Production frontiers and panel data, *Journal of Business and Economic Statistics*, 2(4), 367-374.
- Sen, Amartya K., (1981). Public action and the quality of life in developing countries, *Oxford Bulletin of Economics and Statistics*, 43, 287-319.
- _____, (1988). Sri Lanka's achievements: When and how?, in Srinivasan, T.N., and P.K. Bardhan (eds) *Rural Poverty in South Asia*, New York: Columbia University Press, 549-56.
- Skinner, Jonathan, (1994). What do stochastic frontier cost functions tell us about inefficiency? *Journal of Health Economics* 13, 323-328.
- United Nations Development Programme (UNDP) (1996). *Human Development Report*.

- New York: Oxford University Press.
- Wang, Jia, Dean T. Jamison, Eduard Bos, Alexander Preker, John Peabody (1999).
Measuring Country Performance on Health: Selected Indicators for 115 Countries, Health, Nutrition and Population Series, World Bank.
- Winsten, C.B. (1957). Discussion of Mr. Farrell's paper, *Journal of the Royal Statistical Society A*, 120(3), 282-284.
- World Bank (1993). *World Development Report: Investing in Health*, New York: Oxford University Press.
- _____ (2003). *World Development Indicators*, World Bank, Washington DC.
- World Health Organization (1999). *The World Health Report: Making a Difference*. Geneva: World Health Organization.
- _____ (2000). *The World Health Report: Health Systems. Improving Performance*. Geneva: World Health Organization.
- _____ (2001). *Report of the Scientific Peer Review Group on Health Systems Performance Assessment*, Geneva: World Health Organization.
- http://www.who.int/health-systems-performance/sprg/report_of_sprg_on_hspa.htm